CLAIMS

1. A voice data selector, comprising:

memory means for storing a plurality of voice data
5   expressing voice waveforms;

search means for inputting text information
expressing a text and retrieving voice data expressing a
waveform of a voice unit whose reading is common to that
of a voice unit which constitutes the text from among
10   the voice data; and

selection means for selecting each one of voice
data corresponding to each voice unit which constitutes
the text from among the searched voice data so that a
value obtained by totaling difference of pitches in
15   boundaries of adjacent voice units in the whole text may
become minimum.

2. The voice data selector according to claim 1,
further comprising:
20   speech synthesis means of generating data
expressing synthetic speech by combining selected voice
data mutually.

3. A voice data selection method, the method
25   comprising the steps of:

storing a plurality of voice data expressing voice
waveforms;

inputting text information expressing a text,
retrieving voice data expressing a waveform of a voice
30   unit whose reading is common to that of a voice unit

which constitutes the text from among the voice data; and

selecting each one of voice data corresponding to each voice unit which constitutes the text from among the retrieved voice data so that a value obtained by totaling difference of pitches in boundaries of adjacent voice units in the whole text may become minimum.

4. A program for causing a computer to function as:

memory means for storing a plurality of voice data expressing voice waveforms;

search means for inputting text information expressing a text and retrieving voice data expressing a waveform of a voice unit whose reading is common to that of a voice unit which constitutes the text from among the voice data; and

selection means for selecting each one of voice data corresponding to each voice unit which constitutes the text from among the searched voice data so that a value obtained by totaling difference of pitches in boundaries of adjacent voice units in the whole text may become minimum.

5. A voice selector, comprising:

memory means for storing a plurality of voice data expressing voice waveforms;

prediction means for predicting time series change of pitch of a voice unit by inputting text information expressing a text and performing cadence prediction for

a voice unit which constitutes the text concerned; and

selection means for select from among the voice data the voice data which expresses a waveform of a voice unit whose reading is common to that of a voice unit which constitutes the text, and whose time series change of pitch has the highest correlation with prediction result by the prediction means.

6. The voice selector according to claim 5, wherein the selection means may specify strength of correlation between time series change of pitch of voice data, and result of prediction by the prediction means on the basis of result of regression calculation which performs primary regression between time series change of pitch of a voice unit which voice data expresses, and time series change of pitch of a voice unit in the text whose reading is common to the voice unit concerned.

7. The voice selector according to claim 5, wherein the selection means may specify strength of correlation between time series change of pitch of voice data, and result of prediction by the prediction means on the basis of a correlation coefficient between time series change of pitch of a voice unit which voice data expresses, and time series change of pitch of a voice unit in the text whose reading is common to the voice unit concerned.

8. A voice selector, comprising:

memory means for storing a plurality of voice data

expressing voice waveforms;

prediction means for predicting time length voice unit and time series change of pitch of the voice unit concerned by inputting text information expressing a text and performing cadence prediction for the voice unit in the text concerned; and

selection means for specifying an evaluation value of each voice data expressing a waveform of a voice unit whose reading is common to a voice unit in the text and selecting voice data whose evaluation value expresses the highest evaluation, and in that the evaluation value is obtained from a function of a numerical value which expresses correlation between time series change of pitch of a voice unit which voice data expresses, and prediction result of time series change of pitch of a voice unit in the text whose reading is common to the voice unit concerned, and a function of difference between prediction result of time length of a voice unit which the voice data concerned expresses, and time length of a voice unit in the text whose reading is common to the voice unit concerned.

9.   The voice selector according to claim 8, wherein the numerical value expressing correlation comprises a gradient of a primary function obtained by the primary regression between time series change of pitch of a voice unit which voice data expresses, and time series change of pitch of a voice unit in the text whose reading is common to that of the voice unit concerned.

10. The voice selector according to claim 8, wherein the numerical value expressing correlation comprises an intercept of a primary function obtained by the primary regression between time series change of pitch of a voice unit which voice data expresses, and time series change of pitch of a voice unit in the text whose reading is common to that of the voice unit concerned.

11. The voice selector according to claim 8, wherein the numerical value expressing correlation comprises a correlation coefficient between time series change of pitch of a voice unit which voice data expresses, and prediction result of time series change of pitch of a voice unit in the text whose reading is common to that of the voice unit concerned.

12. The voice selector according to claim 8, wherein the numerical value expressing correlation comprises the maximum value of correlation coefficients between a function which what is given various bit count cyclic shifts to data expressing time series change of pitch of a voice unit which voice data expresses, and a function expressing prediction result of time series change of pitch of a voice unit in the text whose reading is common to that of the voice unit concerned.

13. The voice selector according to any one of claims 5 to 12, wherein the memory means stores phonetic

data expressing reading of voice data with associating it with the voice data concerned; and

wherein the selection means treats voice data, with which phonetic data expressing the reading agreeing with the reading of a voice unit in the text is associated, as voice data expressing a waveform of a voice unit whose reading is common to the voice unit concerned.

14. The voice selector according to any one of claims 5 to 13, wherein further comprising:

speech synthesis means of generating data expressing synthetic speech by combining selected voice data mutually.

15. The voice selector according to claim 14, comprising:

lacked portion synthesis means of synthesizing voice data expressing a waveform of a voice unit in regard to the voice unit, on which the selection means was not able to select voice data, among voice units in the text without using voice data which the memory means stores, and in that the speech synthesis means generates data expressing synthetic speech by combining voice data, which the selection means selected, with voice data which the lacked portion synthesis means synthesizes.

16. A voice selection method, the method comprising the steps of:

storing a plurality of voice data expressing voice

waveforms;

predicting time series change of pitch of a voice unit by inputting text information expressing a text and performing cadence prediction for a voice unit which constitutes the text concerned; and

selecting from among the voice data the voice data which expresses a waveform of a voice unit whose reading is common to that of a voice unit which constitutes the text, and whose time series change of pitch has the highest correlation with prediction result by the prediction means.

17. A voice selection method, the method comprising the steps of:

storing a plurality of voice data expressing voice waveforms;

predicting time length of voice unit and time series change of pitch of the voice unit concerned by inputting text information expressing a text and performing cadence prediction for a voice unit in the text concerned; and

specifying an evaluation value of each voice data expressing a waveform of a voice unit whose reading is common to a voice unit in the text and selecting voice data whose evaluation value expresses the highest evaluation, and in that the evaluation value is obtained from a function of a numerical value which expresses correlation between time series change of pitch of a voice unit which voice data expresses, and prediction result of time series change of pitch of a voice unit in

the text whose reading is common to the voice unit concerned, and a function of difference between prediction result of time length of a voice unit which the voice data concerned expresses, and time length of a voice unit in the text whose reading is common to the voice unit concerned.

18. A program for causing a computer to function as:

memory means for storing a plurality of voice data expressing voice waveforms;

prediction means for predicting time series change of pitch of a voice unit by inputting text information expressing a text and performing cadence prediction for a voice unit which constitutes the text concerned; and

selection means for selecting select from among the voice data voice data which expresses a waveform of a voice unit whose reading is common to that of a voice unit which constitutes the text, and whose time series change of pitch has the highest correlation with prediction result by the prediction means.

19. A program for causing a computer to function as:

memory means for storing a plurality of voice data expressing voice waveforms;

prediction means for predicting time length of a voice unit and time series change of pitch of the voice unit concerned by inputting text information expressing a text and performing cadence prediction for a voice

unit in the text concerned; and

selection means for specifying an evaluation value of each voice data expressing a waveform of a voice unit whose reading is common to a voice unit in the text and

5 selecting voice data whose evaluation value expresses the highest evaluation, and in that the evaluation value is obtained from a function of a numerical value which expresses correlation between time series change of pitch of a voice unit which voice data expresses, and

10 prediction result of time series change of pitch of a voice unit in the text whose reading is common to the voice unit concerned, and a function of difference between prediction result of time length of a voice unit which the voice data concerned expresses, and time

15 length of a voice unit in the text whose reading is common to the voice unit concerned.

20. A voice data selector, comprising:

memory means for storing a plurality of voice data

20 expressing voice waveforms;

text information input means of inputting text information expressing a text;

a search section for searching voice data which has a portion whose reading is common to that of a voice

25 unit in a text which the text information expresses; and

selection means for obtaining an evaluation value according to predetermined evaluation criteria on the basis of relationship between mutually adjacent voice data when each of the searched voice data is connected

30 according to the text which text information expresses,

and selecting combination of voice data, which is outputted, on the basis of the evaluation value concerned.

5   21. The voice data selector according to claim 20, wherein the evaluation criterion is a criterion which determines an evaluation value which shows relationship between mutually adjacent voice data; and

wherein the evaluation value is obtained on the
10   basis of an evaluation expression which contains at least any one of a parameter which shows a feature of voice which the voice data expresses, a parameter which shows a feature of voice obtained by mutually combining voice which the voice data expresses, and a parameter
15   which shows a feature relating to speech time length.

22. The voice data selector according to claim 20, wherein the evaluation criterion is a criterion which determines an evaluation value which shows relationship
20   between mutually adjacent voice data; and that the evaluation value includes a parameter which shows a feature of voice obtained by mutually combining voice which the voice data expresses, and is obtained on the basis of an evaluation expression which contains at
25   least any one of a parameter which shows a feature of voice which the voice data expresses, and a parameter which shows a feature relating to speech time length.

23. The voice data selector according to claim 21
30   or 22, wherein the parameter which shows a feature of

voice obtained by mutually combining voice which the voice data expresses is obtained on the basis of difference between pitches in a boundary of mutually adjacent voice data in the case of selecting at a time

5 one voice data corresponding to each voice unit which constitutes the text from among voice data which expressing waveforms of voice having a portion whose reading is common to that of a voice unit in a text which the text information expresses.

10

24. The voice data selector according to any one of claims 20 to 23, wherein the evaluation criterion further includes a reference which determines an evaluation value which expresses correlation or

15 difference between voice, which voice data expresses, and cadence prediction result of the cadence prediction means; and that the evaluation value is obtained on the basis of a function of a numerical value which expresses correlation between time series change of pitch of a

20 voice unit which voice data expresses, and prediction result of time series change of pitch of a voice unit in the text whose reading is common to the voice unit concerned, and/or a function of difference between prediction result of time length of a voice unit which

25 the voice data concerned expresses, and time length of a voice unit in the text whose reading is common to the voice unit concerned.

25. The voice data selector according to claim 24,

30 wherein the numerical value expressing correlation

comprises a gradient and/or an intercept of a primary function obtained by the primary regression between time series change of pitch of a voice unit which voice data expresses, and time series change of pitch of a voice

5   unit in the text whose reading is common to that of the voice unit concerned.

26. The voice data selector according to claim 24 or 25, wherein the numerical value expressing

10 correlation comprises a correlation coefficient between time series change of pitch of a voice unit which voice data expresses, and prediction result of time series change of pitch of a voice unit in the text whose reading is common to that of the voice unit concerned.

15

27. The voice data selector according to claim 24 or 25, wherein the numerical value expressing correlation comprises the maximum value of correlation coefficients between a function which what is given

20 various bit count cyclic shifts to data expressing time series change of pitch of a voice unit which voice data expresses, and a function expressing prediction result of time series change of pitch of a voice unit in the text whose reading is common to that of the voice unit

25 concerned.

28. The voice selector according to any one of claims 20 to 27, wherein the memory means stores phonetic data expressing reading of voice data with

30 associating it with the voice data concerned; and

wherein the selection means treats voice data, with which phonetic data expressing reading agreeing with reading of a voice unit in the text is associated, as voice data expressing a waveform of a voice unit

5   whose reading is common to the voice unit concerned.

29.   The voice selector according to any one of claims 20 to 28, wherein speech synthesis means of generating data expressing synthetic speech by combining

10   selected voice data mutually.

30.   The voice data selector according to claim 29, comprising:

lacked portion synthesis means for synthesizing

15   voice data expressing a waveform of a voice unit in regard to a voice unit, on which the selection means is not able to select voice data, among voice units in the text without using voice data which the memory means stores, and in that the speech synthesis means generates

20   data expressing synthetic speech by combining a voice data, which the selection means selects, with voice data which the lacked portion synthesis means synthesizes.

31.   A voice data selection method, the method

25   comprising the steps of:

storing a plurality of voice data expressing voice waveforms;

inputting text information expressing a text;

searching voice data which has a portion whose

30   reading is common to that of a voice unit in a text

which the text information expresses;

obtaining an evaluation value according to predetermined evaluation criteria on the basis of relationship between mutually adjacent voice data when each of the searched voice data is connected according to a text which text information expresses; and

selecting combination of voice data, which is outputted, on the basis of the evaluation value concerned.

32. A program for causing a computer to function as:

memory means for storing a plurality of voice data expressing voice waveforms;

text information input means for inputting text information expressing a text;

a search section for searching voice data which has a portion whose reading is common to that of a voice unit in a text which the text information expresses; and

selection means for obtaining an evaluation value according to a predetermined evaluation criterion on the basis of relationship between mutually adjacent voice data when each of the searched voice data is connected according to a text which text information expresses, and selecting combination of voice data, which is outputted, on the basis of the evaluation value concerned.